1. Title

: Weekly stock data for Dow Jones Index

2. Source:

This dataset comprises data reported by the major stock exchanges.

3. Past Usage

This dataset was first used in:

Brown, M. S., Pelosi, M. & Dirska, H. (2013). Dynamic-radius Species-conserving Genetic Algorithm for

the Financial Forecasting of Dow Jones Index Stocks. Machine Learning and Data Mining in Pattern

Recognition, 7988, 27-41.

We request that you provide a citation to this paper when using the dataset. We welcome you to

compare your results against ours in (Brown, Pelosi & Dirska, 2013).

4. Relevant Information

In predicting stock prices you collect data over some period of time - day, week, month, etc.

But you cannot take advantage of data from a time period until the next increment of the time period.

For example, assume you collect data daily. When Monday is over you have all of the data

for that day. However you can invest on Monday, because you don't get the data until the

end of the day. You can use the data from Monday to invest on Tuesday.

In our research each record (row) is data for a week. Each record also has the percentage

of return that stock has in the following week (percent\_change\_next\_weeks\_price). Ideally,

you want to determine which stock will produce the greatest rate of return in the following

week. This can help you train and test your algorithm.

Some of these attributes might not be use used in your research. They were

originally added to our database to perform calculations. (Brown, Pelosi & Dirska, 2013)

used percent\_change\_price, percent\_change\_volume\_over\_last\_wk, days\_to\_next\_dividend,

and percent\_return\_next\_dividend. We left the other attributes in the dataset

in case you wanted to use any of them. Of course what you want to maximize is

percent\_change\_next\_weeks\_price.

Training data vs Test data:

In (Brown, Pelosi & Dirska, 2013) we used quarter 1 (Jan-Mar) data for training and

quarter 2 (Apr-Jun) data for testing.

Interesting data points:

If you use quarter 2 data for testing, you will notice something interesting in

the week ending 5/27/2011 every Dow Jones Index stock lost money.

The Dow Jones Index stocks change over time. The stocks that made up the index in 2011 were:

3M MMM

American Express AXP

Alcoa AA

AT&T T

Bank of America BAC

Boeing BA

Caterpillar CAT

Chevron CVX

Cisco Systems CSCO

Coca-Cola KO

DuPont DD

ExxonMobil XOM

General Electric GE

Hewlett-Packard HPQ

The Home Depot HD

Intel INTC

IBM IBM

Johnson & Johnson JNJ

JPMorgan Chase JPM

Kraft KRFT

McDonald's MCD

Merck MRK

Microsoft MSFT

Pfizer PFE

Procter & Gamble PG

Travelers TRV

United Technologies UTX

Verizon VZ

Wal-Mart WMT

Walt Disney DIS

5. Number of Instances

There are 750 data records. 360 are from the first quarter of the year (Jan to Mar).

390 are from the second quarter of the year (Apr to Jun).

6. Number of Attributes

There are 16 attributes.

7. For each Attribute

*quarter: the yearly quarter (1 = Jan-Mar; 2 = Apr=Jun).*

*stock: the stock symbol (see above)*

date: the last business day of the work (this is typically a Friday)

open: the price of the stock at the beginning of the week

high: the highest price of the stock during the week

low: the lowest price of the stock during the week

close: the price of the stock at the end of the week

volume: the number of shares of stock that traded hands in the week

*percent\_change\_price: the percentage change in price throughout the week*

*percent\_chagne\_volume\_over\_last\_wk: the percentage change in the number of shares of stock that traded hands for this week compared to the previous week*

previous\_weeks\_volume: the number of shares of stock that traded hands in the previous week

next\_weeks\_open: the opening price of the stock in the following week

next\_weeks\_close: the closing price of the stock in the following week

***OBIETTIVO percent\_change\_next\_weeks\_price: the percentage change in price of the stock in the following week***

*days\_to\_next\_dividend: the number of days until the next dividend*

*percent\_return\_next\_dividend: the percentage of return on the next dividend*

8. Missing Attribute Values:

None